Realistic Broadcasting Using Multi-modal Immersive Media

Sung-Yeol Kim, Seung-Uk Yoon, and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST),
1 Oryong-dong, Buk-gu, Gwangju, 500-712, Korea
{sykim75, suyoon, hoyo}@gist.ac.kr

Abstract. Realistic broadcasting is considered as a next generation broadcasting system supporting user-friendly interactions. In this paper, we define multi-modal immersive media and introduce technologies for a realistic broadcasting system, which are developed at Realistic Broadcasting Research Center (RBRC) in Korea. In order to generate three-dimensional (3-D) scenes, we acquire immersive media using a depth-based camera or multi-view cameras. After converting the immersive media into broadcasting contents, we send the immersive contents to the clients using high-speed and high-capacity transmission techniques. Finally, we can experience realistic broadcasting represented by the 3-D display, 3-D sound, and haptic interaction. Demonstrations show two types of broadcasting systems: the system using a depth-based camera and the system using multi-view cameras. From the realistic broadcasting system, we can generate new paradigms for the next generation digital broadcasting.

Keywords: Realistic broadcasting, Immersive media, Technologies of a realistic broadcasting system.

1 Introduction

As the rapid development of telecommunication techniques and high-speed networks, we live in an age of the information revolution and the digital epoch. Humans acquire useful knowledge and information through the Internet since high-speed networks are connected with high-performance personal computers. We cannot only feel deep impressions by a high definition television with a large screen and a high power speaker, but also call to someone using a cellular phone with moving pictures. In addition, banking services and product purchases are possible at home. The digital technologies make a human life more convenient and livelier.

It is not too much to say that the essence of the digital age is the multimedia technologies for digital broadcasting. The digital broadcasting system converts analog multimedia data into digital multimedia data and then transmits the digitized data to the end users. The digital broadcasting is suitable for a high-quality and multi-channel broadcasting in comparison with an analog broadcasting. Furthermore, a digital broadcasting system can make use of the frequency bandwidth effectively and have great advantages for the data broadcasting.

Y.-S. Ho and H.J. Kim (Eds.): PCM 2005, Part II, LNCS 3768, pp. 164 – 175, 2005.

Figure 1 shows the whole development trends of the digital broadcasting [1]. As shown in Fig. 1, the tendency for broadcasting services will change from one-directional services to bi-directional services or interactive services such as a stereoscopic TV, a three-dimensional (3-D) TV, and a realistic broadcasting. In the broadcasting service quality aspect, the next generation broadcasting will be a high quality 3-D broadcasting system.

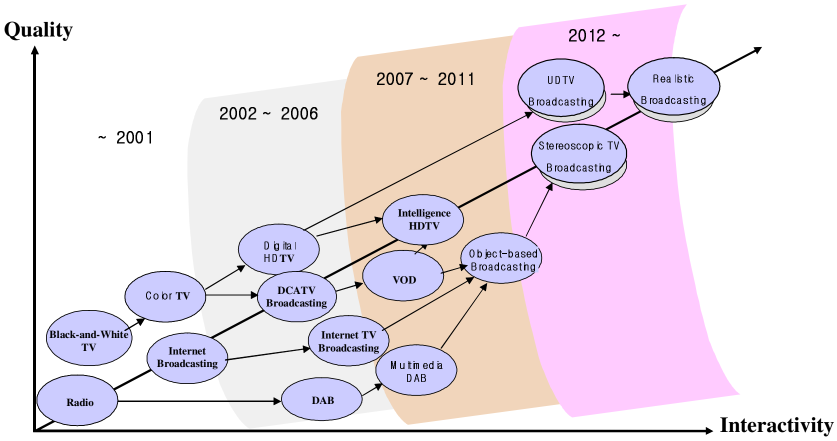

Fig. 1. The trends in digital broadcasting

Especially, realistic broadcasting makes an appearance for the next generation broadcasting. Realistic broadcasting provides not only high quality visual services, but also a variety of user-friendly interactions. Unlike other digital broadcasting systems, we can experience the realism through our five senses.

Several countries have already served 3-D broadcasting experimentally using their satellites. Furthermore, a large number of researchers are interested in developing 3-D displays and 3-D data processing techniques since the scale for 3-D data markets is supposed to be three billion dollars in 2010. In this paper, we will survey the technologies for a realistic broadcasting system, described by Realistic Broadcasting Research Center (RBRC) in Korea, so as to keep pace with this trend.

2 Immersive Media and Realistic Broadcasting

The existing media services focused on 2-D audio-visual data owing to the technological limits. The audio-visual data were not enough to give us the vividness through the five senses of human. However, we can represent and treat high-capacity media easily, as the computer hardware and software progress rapidly.

It is conservative to say that enhanced digital multimedia services lie in the advent of multi-modal immersive media and the technology of realistic broadcasting. Multi-modal immersive media indicate the data overcoming the spatial-temporal limit. As examples of the multi-modal immersive media, there are computer graphic models, 3-D video, multi-channel audio, haptic data, and the information for the sense of smell.

Realistic broadcasting can be defined by a broadcasting service system using multi-modal immersive media so as to endow the end users with the realism. Immersive contents indicate the broadcasting contents with the multi-modal immersive media. Figure 2 shows the conceptual illustration for realistic broadcasting.

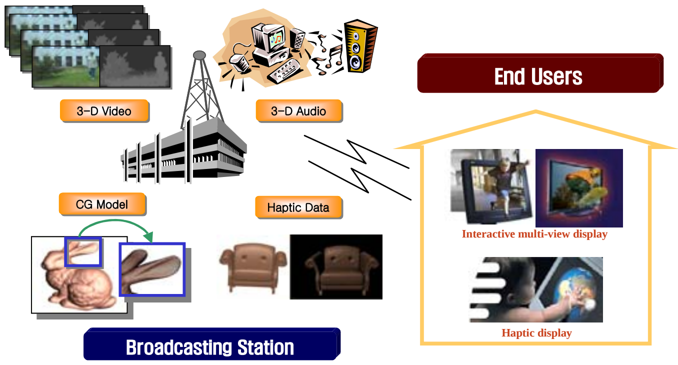

Fig. 2. Realistic broadcasting services

A number of international research centers and universities have persevered in their efforts for developing the core technologies related to realistic broadcasting. Holographic display techniques have been developed in Massachusetts Institute of Technology. The ATTEST project, archived from 2002 to 2004, gave us a possibility for realizing 3-D TV [2]. With a new project named by 3DTV, started in September 2004, the development of the commercial European 3-D broadcasting system has been in progress [3]. O'Modhrain and Oakley [4, 5] discussed the potential role that haptic or touch feedback might play in supporting a greater sense of immersion in the next generation broadcasting system [6]. In recent years, multi-view video coding has been a big issue in MPEG 3DAV standards [7]. Likewise, NHK research center in Japan have been developing 3-D display and 3-D sound technologies [8], and Fraunhofer Heinrich Hertz Institute (FhG-HHI) in Germany exploits the efficient compression scheme based on 3-D object modeling [9].

Previous researches related to realistic broadcasting have been archived by the unit of small group individually. In order to develop a feasible realistic broadcasting, we need to integrate various technologies. In this paper, we will introduce the incorporated technologies to realize a realistic broadcasting system.

3 Technologies for Realistic Broadcasting

The realistic broadcasting system can be divided into four parts largely; the acquisition of immersive media, the editing of immersive contents, the transmission of immersive contents, and the display of immersive media. Figure 3 shows the key technologies of the realistic broadcasting system.

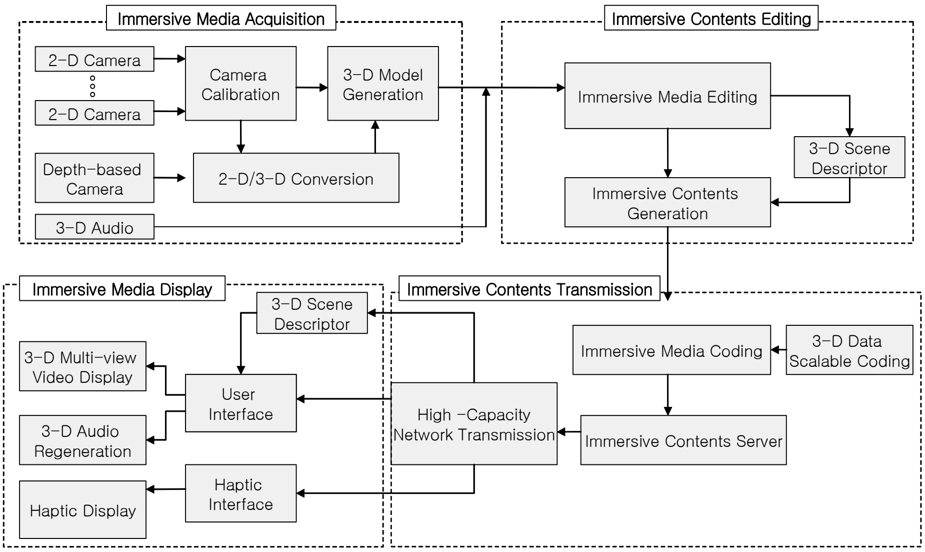

Fig. 3. Overall system of realistic broadcasting

In the process of the acquisition for immersive media, we generate 3-D models and 3-D photo-realistic backgrounds using a number of 2-D stereo cameras or a depth-based camera. Also, we can obtain 3-D sound corresponding to the 3-D models. In general, we pass through a camera calibration so as to obtain the exact 3-D data.

In order to make 3-D scenes, called by the immersive contents, from the acquired 3-D data, we need to edit the immersive media. In the process of the editing for immersive contents, we combine the 3-D video, 3-D audio, and other additional data, such as haptic information.

Basically, the immersive contents require a tremendous amount of data. In order to store the immersive contents compactly and transmit them efficiently, we need to compress the immersive contents. Moreover, we also need to develop a proper network protocol for the high-capacity transmission.

Finally, we display received immersive media with a high definition television. At the client side, we can watch 3-D scenes with stereoscopic displays. Moreover, various user-friendly interactions are supported by a view interaction, a haptic display, and bi- directional data transmission.

3.1 Acquisition of Immersive Media

Before generating immersive contents for realistic broadcasting, we need to acquire immersive media from 3-D data acquisition devices. After obtaining immersive data, we need to do multi-modal immersive media modeling. Figure 4 shows the immersive media acquisition technologies.

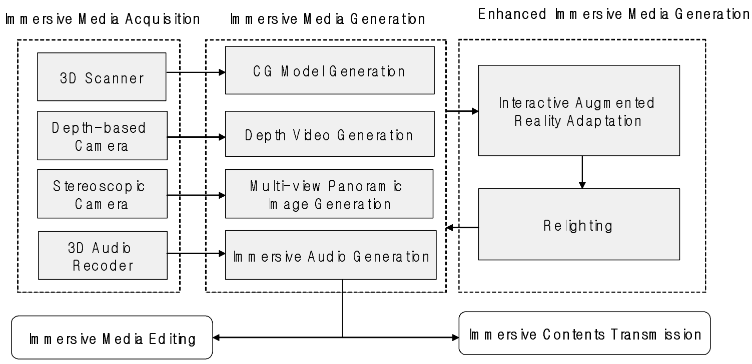

Fig. 4. Immersive media acquisition

3.1.1 Generation of Computer Graphics Models

In order to acquire computer graphics (CG) models, we use 3-D data scanning device or graphic tools. Basically, CG models are composed of 3-D coordinate data and attribute data such as colors, normal vectors, and textures. We need to estimate and compensate the colors and surface materials of CG models since the acquisition environments are various according to lighting and temperature conditions. Figure 5 shows the devices for the immersive media acquisition including a 3-D scanner.

(a) 3-D scanner (b) Multi-view Camera (c) Depth-based Camera

Fig. 5. Immersive media acquisition devices

3.1.2 Generation of Depth Video

Depth video indicates a sequence of depth images and texture images. For accomplishing the realistic broadcasting system, we should generate 3-D scenes in real-time. With a depth-based camera and a number of stereo cameras, we can generate 3-D scenes based on depth video. While we can obtain the depth data through stereo matching when we use a set of stereo cameras, we can get the depth information directly by the depth-based camera.

3.1.3 Generation of Multi-view Panoramic Image

Multi-view panoramic images are used for 3-D backgrounds in our realistic broadcasting system. After obtaining images by a set of stereo cameras, we execute the local stitching and the global stitching so as to make a 3-D panoramic image. Finally, we should optimize the geometry data since the stitching can cause errors.

3.1.4 Relighting

When we combine 3-D models and photo-realistic environment, we need to know the similar light conditions with photo-realistic environment to render 3-D scenes naturally. As shown in Fig. 6, after obtaining a high dynamic range (HDR) image by estimating the lighting conditions, light probe, and high-resolution cameras, we adapt the acquired light conditions to 3-D scenes.

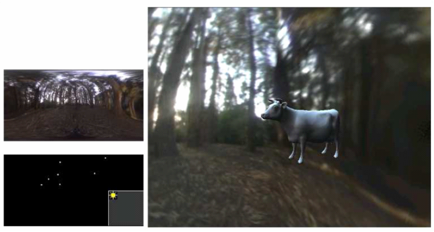

Fig. 6. Relighting using HDR

3.1.4 Generation of Immersive Audio

In general, 3-D video data can play a role for realistic broadcasting contents when 3-D audio and 3-D sound are accompanied. Therefore, the generation of immersive audio is an important part in the generation of immersive media. The immersive audio and sound are obtained by evaluating the audio direction and distance using a head related transfer function (HRTF) estimator.

3.2 Editing of Immersive Contents

After acquiring immersive media, we need to edit the immersive media and convert them into immersive contents for realistic broadcasting. As shown in Fig. 7, there are an auto-segmentation, special effects, an editing system for immersive media, and a broadcasting item generation.

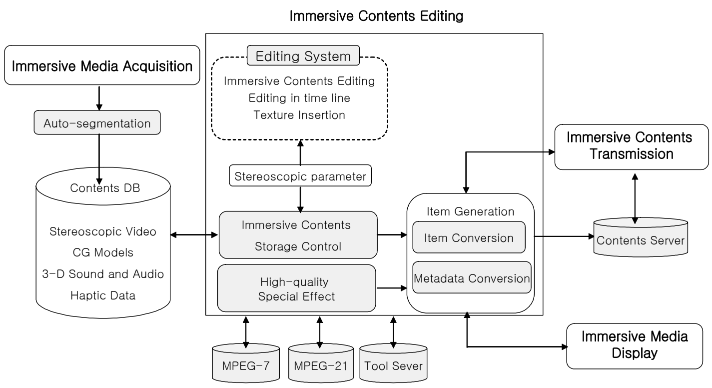

Fig. 7. Immersive Contents Editing

3.2.1 Auto-segmentation and Special Effects

As the essential techniques for immersive content editing, auto-segmentation defines immersive media as a unit of object and extracts the objects from 3-D backgrounds. After the segmented objects are applied by special effects such as warping and morphing techniques according to usages, new immersive contents are generated with combination between the segmented objects and other immersive media.

3.2.2 Editing System

In order to synthesize each immersive media and generate 3-D scenes in timeline, we need to obtain a 3-D scene descriptor. The 3-D scene descriptor indicates the relationship between CG models and generated depth video in each frame. In an editing system for realistic broadcasting, we reallocate the immersive media by 3-D scene editors and generate the 3-D scene descriptor. At the client side, we can regenerate 3-D scenes by analyzing the transmitted 3-D scene descriptor.

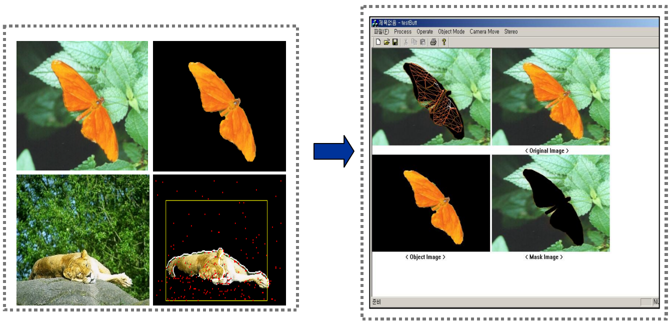

Fig. 8. Auto-segmentation and editing system

3.2.3 Broadcasting Item Generation

The broadcasting item generation is important parts to support an interactive data broadcasting system. In order to communicate broadcasting servers and clients, we should provide a variety of items and metadata. We can generate and convert items for realistic broadcasting by adapting MPEG-21 digital item adaptation (DIA) and stereoscopic techniques in the MPEG-7 standard.

3.3 Transmission of Immersive Contents

For realistic broadcasting services, we need to compress the immersive contents efficiently and transmit them through high-speed networks. Immersive contents transmission technology supports multi-view video coding, CG model compression, immersive media scalability, high-capacity transmission, and immersive content server technology. Figure 9 shows the immersive contents transmission technologies.

3.3.1 Graphic-Based Model Coding

We need to compress graphic-based models since they also require a large amount of data. Graphic-based model coding can be divided into two parts: static model coding and dynamic model coding. For static model coding, we usually use a mesh-based compression scheme provided by the MPEG-4 SHNC standard. Also, we mainly use an interpolator compression (IC) scheme for dynamic models.

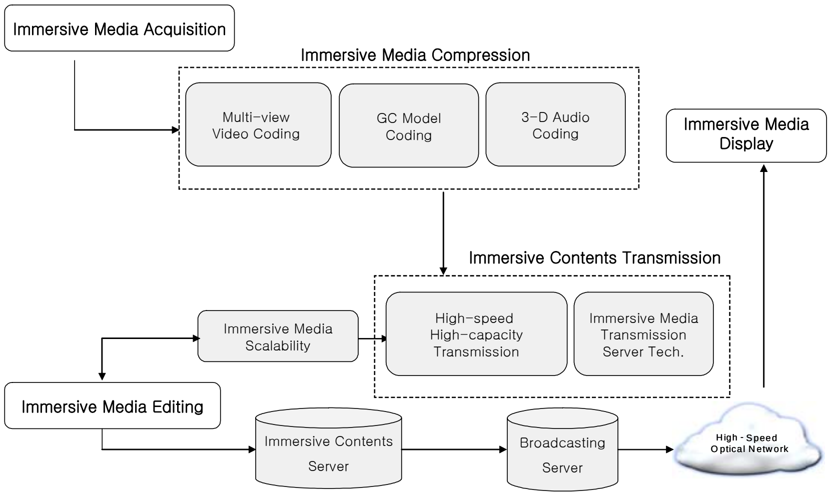

Fig. 9. Immersive contents transmission

3.3.2 Multi-view Video Coding

Realistic broadcasting should provide natural and continuous 3-D video in real-time. In order to support such a multi-view 3-D video, it is beneficial for us to use the 3-D multi-view video. Basically, we develop the multi-view video coding algorithms to transmit the video data within limited-bandwidth networks. In order to compress the multi-view video, we use intermediate view reconstruction (IVR) [10] and layered depth image (LDI) [11].

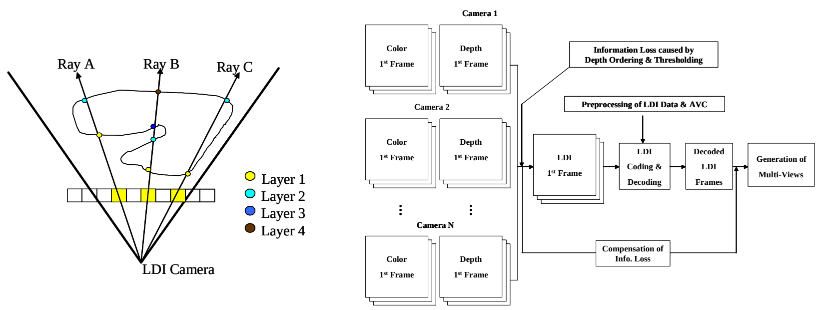

Fig. 10. Multi-view video compression using LDI

3.3.3 Immersive Media Scalability

Immersive media need to be scalable so as to serve the realistic broadcasting according to the capacity and environment of clients. By using the scalability of the immersive media, we can enjoy the realistic broadcasting in anyway and anywhere. We support the immersive media scalability according to the region of interest, view-dependence, and the distance of immersive media.

3.2.4 High-Speed and High-Capacity Transmission

In order to send the immersive content with high speed, we need to change the amount of transmitted data according to the network situation and protocols. We need transcoding schemes and layered representation techniques for immersive media and we consider the robustness and required bandwidth. Furthermore, a priority scheme for determining the order of packets is required for a high-capacity transmission.

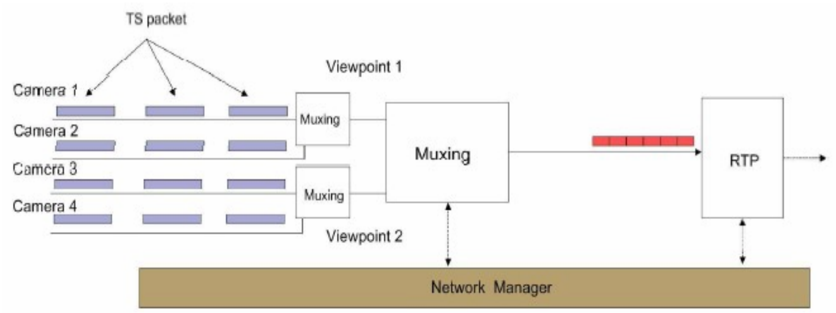

Fig. 11. Immersive contents transmission

3.2.5 Immersive Contents Transmission Server Technology

Since the clients share immersive contents simultaneously in the realistic broadcasting, it is necessary to develop the transmission server technology for supporting the interactive data broadcasting. Moreover, we need to exploit the packet distribution techniques with immersive contents. The content transmission server technologies are composed of server composition and interactive support techniques.

3.4 Display of Immersive Media

The immersive media display technology provides 3-D display, 3-D sound and haptic display. There are surrounding environment recognition services, stereoscopic format conversions, haptic display techniques, and 3-D audio regeneration techniques in immersive media display technologies. Figure 7 shows the immersive media display technologies.

3.4.1 Surrounding Environment Recognition

The surrounding environment recognition is used to trace the eyes of clients. As analyzing the viewpoints and actions of clients, terminals display various stereoscopic images with respect to the viewpoints of the end users. We develop the surrounding environment recognition technique using a head-tacking camera.

3.4.2 Stereoscopic Format Conversions

In order to adapt the existing immersive contents to terminals, we need to develop the 3-D converting techniques. When the input contents are provided by page-flip, interlace, interleave, top-down, side-by-side, and sync double schemes, we convert the input contents into a proper format that is possible to display at a terminal.

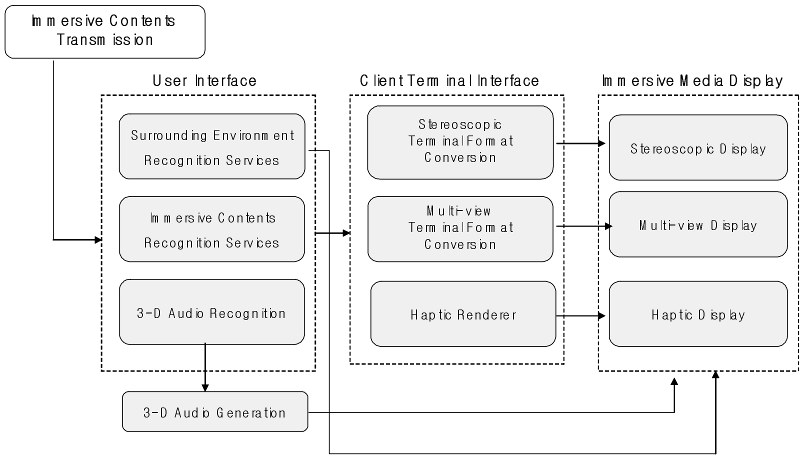

Fig. 12. Immersive media display

3.4.3 Haptic Display

The realistic broadcasting can not only watch and listen to the broadcasting contents by 3-D display and 3-D sound, but also support haptic display so as to touch and control the immersive contents. Haptic information is a sense with human's muscles and skins. We can use the haptic data in educational and entertainment programs such as a scientific documentary and a fishing television program. Also, haptic information can be used for a home shopping channel by supporting surface and shape information of the goods.

3.4.4 Audio Recognition and 3-D Audio Generation

We need to develop an interface using an audio recognition technique for the realistic broadcasting system. We can get useful information from a television by an audio interface, as well as changing a channel and immersive contents. In addition, the acquired immersive audio and sound can be regenerated by 3-D sound techniques.

4 Demonstrations of Realistic Broadcasting

We have evaluated the system using multi-view stereo cameras [12] and the system using a depth-based camera [13] for the realistic broadcasting in the forum provided

Fig. 13. System using multi-view cameras

by Information Technology Research Center (ITRC) in 2005. The system using multi-view cameras includes the surrounding recognition, high-capacity and high-speed transmission, and stereoscopic conversion techniques. On the other hand, the system using a depth-based camera includes the 3-D data modeling, stereoscopic display and haptic rendering techniques.

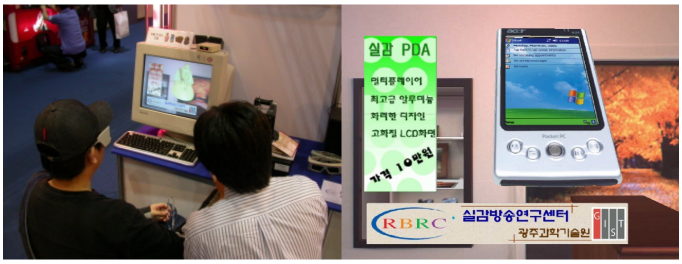

Fig. 14. System using a depth-based camera

5 Conclusions

In this paper, we define a realistic broadcasting system and review the technologies for realistic broadcasting, described in RBRC. In order to generate immersive media, we use the multi-view cameras and a depth-based camera. After converting the immersive media into immersive contents for the 3-D scenes, we send the immersive contents to the client by high-speed and high-capacity transmission techniques. Finally, we can experience the 3-D display, 3-D sound, and haptic display at the client side. The trends of recent broadcasting are changing from one-directional broadcasting systems to interactive bi-directional broadcasting systems. The realistic broadcasting, supporting the interactive bi-directional broadcasting system, is supposed to be the next revolution in the history of television.

Acknowledgements. This work was supported in part by Gwangju Institute of Science and Technology (GIST), in part by the Ministry of Information and Communication (MIC) through the Realistic Broadcasting Research Center (RBRC), and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

References

1. Ministry of Science and Technology of Korea, National Technology Roadmap, 2003.
2. Redert, A., Op de Beeck, M., Fehn, C., IJsselsteijn, W., Pollefeys, M., Van Gool, L., Ofek, E., Sexton, I., Surman, P.: ATTEST–Advanced Three-Dimensional Television System Technologies. Proceeding of International Symposium on 3D Data Processing (2002) 313–319.
3. http://3dtv.zcu.cz/.
4. O'Modhrain, S., Oakley, I.: Touch TV: Adding Feeling to Broadcast Media. Proceeding of the European Conference on Interactive Television (2003) 41-47.

5. O'Modhrain, S., Oakley, I.: Haptic Interfaces for Virtual Environment and Teleoperator Systems. Proceedings of Haptics (2004) 293-294.
6. Cha, J., Ryu, J., Kim, S., Eom, S., Ahn, B.: Haptic Interaction in Realistic Multimedia Broadcasting. Proceeding of Pacific Rim Conference (2004) 482-490.
7. Multi-view Video Coding using Shared Picture Memory and Shared Vector Memory. ISO/IEC JTC1/SC29/WG11 M11570 (2005).
8. McMillan, L.: A List-Priority Rendering Algorithm for Redisplaying Projected Surfaces. University of North Carolina, (1995) 95–105.
9. Duan, J., Li, J.: Compression of the LDI. IEEE Transaction on Image Processing, Vol. 12, No. 3, (2003) 365–372.
10. Bae, J. W., Park, H. J., Kim, E. S., Yoo, J.: An Efficient Disparity Estimation Algorithm Based on Spatial Correlation. Journal of SPIE Optical Engineering Vol.42, No.1, (2003) 176-181.
11. Ho, Y.S., Yoon, S.U., Kim, S.Y., Lee, E.K.: Preliminary Results for Multi-view Video Coding using Layered Depth Image. ISO/IEC JTC1/SC29/WG11 M11916 (2005).
12. Lee, S., Lee, K., Han, C., Yoo, J., Kim, M., Kim, J.W.: Multi-view Stereoscopic High-Definition Video over IP Networks for Next Generation Broadcasting. Proceeding of Pacific Rim Conference (2005).
13. Cha, J., Kim, S., Kim, S.Y., Kim, S., Yoon, S.U., Oakley, I., Ryu, J., Lee, K. H., Woo, W., Ho, Y.S.: Client System for Realistic Broadcasting: A First Prototype. Proceeding of Pacific Rim Conference (2005).